A Deep Learning Model of Prescient Ideas Demonstrates that they Emerge from the Periphery

Paul Vicinanza[1], Amir Goldberg[1], and Sameer B. Srivastava[2]

[1]Stanford Graduate School of Business
[2]University of California Berkeley, Haas School of Business

Forthcoming, *PNAS NEXUS*

Abstract

Where do prescient ideas—those that initially challenge conventional assumptions but later achieve widespread acceptance—come from? Although their outcomes in the form of technical innovation are readily observed, the underlying ideas that eventually change the world are often obscured. Here we develop a novel method that uses deep learning to unearth the markers of prescient ideas from the language used by individuals and groups. Our language-based measure identifies prescient actors and documents that prevailing methods would fail to detect. Applying our model to corpora spanning the disparate worlds of politics, law, and business, we demonstrate that it reliably detects prescient ideas in each domain. Moreover, counter to many prevailing intuitions, prescient ideas emanate from each domain's periphery rather than its core. These findings suggest that the propensity to generate far-sighted ideas may be as much a property of contexts as of individuals.

Introduction

Where do prescient ideas—those that initially challenge conventional assumptions but later achieve widespread acceptance—come from? Prior research, predominately focused on the scientific and technical realm, traces these ideas to brilliant individuals[1,2] or the characteristics of inventors and their teams.[3,4,5,6,7] Because ideas themselves are difficult to observe, this literature focuses on how they manifest in the form of tangible artifacts such as patents or scientific publications. Yet most ideas, from disruptive business strategies to paradigm-shifting legal interpretations, do not translate into discrete artifacts that are amenable to such analysis. Instead, they fundamentally transform the prevailing assumptions and beliefs in a domain—but often in subtle ways.

In the realm of politics, for instance, legislators regularly introduce and contest novel ideas that later become taken-for-granted assumptions. In the debates that raged about civil rights legislation in the U.S. during the 1960s, two of the staunchest opponents of these bills were Senators John Stennis and James Eastland. Although both voted against every major piece of civil rights legislation, they diverged in the nature of the opposition they put forward. Whereas Eastland—the least prescient senator according to our model—framed his opposition using overtly racist arguments (see SI and Fig. 1), Stennis, the senator deemed most prescient by our model, was among the first to base his objections on the principles of "color blindness," limited government, and individual freedom.[8] This more muted and indirect set of arguments would later become commonplace among opponents of civil rights legislation, laying the groundwork for contemporary conservative talking points on race relations in the U.S. Stennis' strategy for reframing race relations in the United States cannot be simply reduced to a single statement or artifact. Nor was it merely about novel rhetoric. Instead, it involved discourse that sought to deftly rearrange the tacit assumptions about the interrelationships between race, policy, and conservative values.

---*Insert Figure 1 Linewidth Here*---

Building on this intuition, we propose that prescient ideas are incommensurable, in the Kuhnian sense,[9] with conventional logic. Kuhn, in his groundbreaking work *The Structure of Scientific Revolution*, argued that measuring paradigm shifts requires "indirect and behavioral evidence" and suggested that a promising source of such evidence is the language scientists use to describe the world. He gives the example of the word "planet" during the Copernican Revolution. By stripping the sun of the designation of “planet,'” Copernicans reconceptualized all celestial bodies and the relationships between them. Thus, shifts in language can sometimes signal deeper changes in cognitive understanding among actors in a domain, leaving linguistic markers---such as the use of “planet” in a somewhat modified sense---that foretell fundamental changes in the world.

Groundbreaking advances in natural language processing and deep learning have made it possible to extract such markers from the natural language that people use. Word embedding models capture culturally salient dimensions by representing words in a high-dimensional space. Contextual embedding models, which explicitly account for the context of each word, allow these dimensions to vary across contexts.[10,11,12,13] We identify the linguistic markers of prescient ideas using Bidirectional Encoder Representations from Transformers (BERT),[14] a deep neural network that encodes the semantic and contextual information of language. Taking advantage of BERT’s ability to predict words given their context, we first develop a measure of *contextual novelty*: utterances that are poorly predicted by the model. To measure *prescience,* we compare the contextual novelty at the time ideas are first expressed within a domain to contextual novelty at a later point in time. Ideas that are incommensurable with conventional logic at the time of their expression but become more commensurable in the future are deemed prescient by our model. Importantly, this approach is agnostic to whether prescient individuals or groups are themselves the inventor of an idea, are early adopters, or are intentional prognosticators.

Our novel method enables us to unearth prescient ideas from natural language independent of the form in which the idea is ultimately realized and across a broad range of domains. Moreover, it allows us to answer a fundamental question in the science of ideas and innovation: Where do prescient ideas come from? Existing theories and evidence lead to competing expectations about whether prescient ideas will emerge from the periphery or the core of a given domain. On one hand, prior work on technological innovation suggests that transformative ideas emerge from the periphery because actors on the outskirts are less bound by institutional constraints and have greater freedom to explore new ideas.[15,16,17] Yet in many non-technical domains such as law, far-sighted ideas are commonly assumed to emerge from sage and established actors such as Supreme Court justices.[18] Such actors are believed to have the resources and experience needed to rethink prevailing assumptions and the power to implement their vision. Even in the realm of technological innovation, prescient ideas often seem to emanate from the core, rather than peripheral, positions. Apple’s groundbreaking iPhone, for example, was released when it was already among the world’s most dominant technology companies. We attribute these mixed findings and contradictory intuitions to the fact that prior work has focused on studying the tangible artifacts that result from prescient ideas rather than interrogating where the ideas themselves came from.

To address this gap, we apply our method for detecting the linguistic markers of prescient ideas to three corpora that span the disparate domains of politics, law, and business. We first validate our measure by establishing that prescience is recognized and rewarded in a given field: Prescient court decisions are more highly cited, prescient politicians ascend to more powerful committees, and prescient firms have higher future market valuations. Delving into the origins of prescient ideas, we next demonstrate that highly prescient ideas emanate from the periphery rather than the center of the field.

Defining Prescient Ideas

What makes an idea prescient? One key ingredient would appear to be novelty, as prescient ideas depart from the accepted conventions of the moment. Yet abundant prior research has shown that novelty does not by itself guarantee success.[19,20,21] We propose that prescient ideas have two essential properties. First, they are novel in a particular way: they rethink the contextual assumptions that predominate a given field. By contextual assumptions, we mean those that: (a) are central to a domain's logics of action, and (b) guide a set of interdependent choices about how to configure activities for success in the field. In 1970, for example, Congress passed the Racketeer Influenced and Corrupt Organizations Act (RICO) to target organized criminal enterprises—in particular the Mafia. A small group of imaginative prosecutors soon seized upon RICO's ambiguous language to prosecute such wide-ranging civil crimes as mail fraud and stock manipulation.[22] This approach was contextually novel in that it applied existing statutes intended for one set of actors to an entirely different class of actors and criminal activities.

Second, prescient ideas foreshadow how the domain will evolve in the future. While RICO's scope was initially limited, this approach to interpreting and applying RICO statutes well beyond their original scope became commonplace among prosecutors and judges. It has since been used to prosecute organizations ranging from the Catholic Church to Major League Baseball to British Petroleum (BP) following the Deepwater Horizon oil spill.[22] Notice that these two ingredients—contextual novelty when an idea is initially articulated and widespread acceptance in its future—

can appear in an individual’s discourse even when the person does not explicitly set out to predict the future, influence others, or even change the world. Moreover, prescient ideas can only be detected after the fact—that is, once the future state of the world is known.

Developing a Language-based Measure of Prescience

Ideas are, of course, hatched by individuals and often expressed in discourse. The core idea of our approach is to measure the extent to which ideas expressed in routine discourse are linguistic markers of prescient ideas. Our technique relies on the intuition that prescient ideas depart from prevailing ideas at the time of introduction but become commonplace in the future.[23,5] In particular, we use BERT, which learns the semantic and syntactic structure of language (in part) through a masked-word prediction task.[14] BERT repeatedly predicts different masked (hidden) words in a sentence given the rest of the sentence, with the aim of minimizing the cross-entropy loss between the predicted and actual word. While most researchers apply BERT’s model architecture to solve downstream tasks such as machine translation or text classification, we use the probabilistic features of the model to assess the extent to which a given set of ideas are prescient in their field.

Similar to how prior work trains separate word embedding models on a temporally split corpus to uncover semantic shifts in word meaning,[10] by training separate BERT models over a split corpus, we reveal how the likelihood of specific words, phrases, and sentences evolves over time. Following standard practice, we begin with a pre-trained model and then fine-tune it to a given time interval (e.g., a year or presidential term) in each of our domains of interest.

---Insert Figure 2 Page Width Here---

To measure prescience, we begin by considering perplexity, the exponentiated cross-entropy loss, which can be intuitively understood as the inverse-likelihood of the model generating a word or a document (normalized by the number of words). Higher perplexity scores correspond to unusual or unexpected utterances. We define *contextual novelty*, *CN*(*s*), as the product of word-level perplexities in a sentence, *s*, normalized by the number of words in the sentence. We use mean sentence-level perplexity values to derive a measure of contextual novelty at the document level. *Prescience* can then be operationalized as the percentage decrease in *contextual novelty* of a sentence between two time periods. Fig. 2 provides a schematic representation of our measurement approach, using a sentence from our legal data that was deemed highly prescient by our model. Depending on the context, we can then aggregate our measure of prescient discourse at different levels of analysis. In some settings, we can identify individuals who are apt to express prescient ideas. These individual-level measures can be aggregated to the level of social groups or organizations that might be more salient in other contexts. In still other settings, the relevant unit of analysis might be a prescient document.

This approach to measuring prescience offers several advantages over prior work such as topic models, n-grams, or TF-IDF vectorization.[24,23,5] When applied to scientific innovation, these methods are commonly used to identify the introduction of new terms or bigrams. This assumes that novel ideas necessarily translate into new terminology. Yet innovative ideas in non-technical domains, such as those studied in this paper, are rarely accompanied by new terms. Instead, prescient ideas emerge as actors reconceptualize existing topics in incommensurate ways. Both Stennis and Eastland ardently denounced racial desegregation, yet only Stennis is considered prescient by framing the issue around color-blindness and limited government. By taking context into account, contextual embeddings capture such nuance.

Another feature of our approach is that any potential biases toward high perplexity sentences—such as rare tokens or errors in optical character recognition—are netted out in the numerator. Likewise, discussions unrelated to prescience are netted out because the likelihood of a sentence must shift over time to result in a non-zero contribution to prescience. Traditional pre-processing steps, such as removing stop words and punctuation, stemming, or converting to lowercase, are unnecessary in contextual embedding models. Unlike topic models, our approach does not require tuning hyperparameters—though the researcher does have to make choices about how to partition the data. For more details, please refer to the SI.

Results

Empirical Settings

We apply this method to identify prescient ideas in the domains of politics (4.9 million floor speeches given by members of the U.S. House of Representatives and the U.S. Senate), law (4.2 million rulings on U.S. State and Federal cases), and business (108,334 quarterly earnings calls in which the management teams of publicly traded firms lay out their vision and strategy for the company to financial analysts who cover their stock). Given that the corpora vary considerably in the time periods they cover and the nature of the discourse they include, we use slightly different approaches to fine-tuning BERT and defining the salient time horizon across the three (SI Appendix).

Model Validation

First, we return to the case studies of the two senators who were identified by our model as most and least prescient in the political dataset (Figure 1): Senators John Stennis and James Eastland, respectively. While both Stennis and Eastland were segregationists from Mississippi, Stennis is remembered as a conscientious leader of the Senate, while Eastland now symbolizes southern intransigence on racial integration. Scholars have traced these divergent judgments to Stennis' pioneering use of novel arguments—such as appeals to "color blindness," limited government, and individual freedom—to oppose civil rights legislation.[8] Our model echoes this conclusion, rating Stennis' 1964 utterance, "I will join hands with any Senator in trying to devise a voluntary, noncoercive, and noncompulsive plan under which both races can work and live together in harmony and make progress," as being in the 99th percentile in prescience. Our method also correctly anticipates the impending disappearance of Eastland's overt brand of bigotry: his 1963 statement against civil rights legislation, "can he rationally write that races are equally law-abiding when the police statistics jeer at the thought?" is more than two standard deviations below mean prescience. Other senators in the most/least prescient list add further credence to our measure. Russell B. Long, the second most prescient Senator, chaired the Senate Finance Committee for over 25 years and is widely regarded as one of the most powerful, respected, and influential Senators of the 20th century.[25] In contrast the second least prescient senator, Jesse Helms, ardently opposed non-white, disability, gay, and feminist rights and is perhaps the most conservative politician of the modern era.

Table 1: Most and Least Prescient uses of the Word "Race"

Most Prescient	I am astonished at the crude attempt of the President to inject a race issue into what is essentially a matter of public policy.
	This is an attempt to exploit Mr. Weavers name and race to gain a purely partisan end.
	The issue was a race issue pure and simple.
	Let us take the necessary legislative action to leave no room for doubt that we will no longer tolerate in the Capital of the United States discrimination in real estate transactions based upon a race issue.

	The new Committee on Equal Employment Opportunity is moving to enforce Executive orders against race discrimination in hiring workers on work done under Government contract
Least Prescient	So my friends the white race is on trial very seriously on trial.
	Let me say to the gentleman in my familiarity with his record and the legislation he has introduced there has never been anything that was for the benefit of the white race.
	There is no race suicide in the great State of Michigan.
	Already the conflict in Algeria has degenerated into race war.
	I think I hardly need to add that many prominent and distinguished members of the white race are members of the organization on its board of directors and in key positions on its various committees.

The most and least prescient uses of the word “race” by politicians in the 87th Congress (1961-1963). Examples are restricted to the context of racial groups—that is, excluding sentences discussing such phrases as “space race” or “political race.”

To illustrate the granularity and efficacy of our methodological approach, we examine the most and least prescient uses of the word “race” in Table 1. Contextual embeddings compute perplexities at the level of individual words, and we can examine instances when the word “race” is used in ways that the model deems highly prescient (i.e., the focal word is better predicted by a model trained on text from future periods relative to a model trained on the contemporaneous period) versus minimally prescient. Conducting this analysis for text from the 87th Congress (1961-1963), we observe stark differences between how the term is used in its most versus least prescient forms. Highly prescient uses are uttered in the context of discourse that seeks to eradicate racial discrimination and that denounces race-bating argumentation. The least prescient uses seem strikingly anachronistic by today’s standards—for example, using terminology such as the “distinguished members of the white race” and “race suicide.”

---Insert Figure 3 Line Width Here---

To more systematically validate our measurement strategy, we conduct an additional analysis based on the so-called “Gingrich Senators,” a group of 33 Republican members of the U.S. House of Representatives who, according to scholarship in political science,[26] served with Newt Gingrich in the House and espoused a new style of highly partisan, divisive, and obstructionist politics. These Representatives were later elected to the U.S. Senate, to which they brought an overtly antagonistic style of discourse. The Gingrich Senators represent a textbook example of prescience: They were a group of individuals who used language that was indicative of a new brand of politics, which later became institutionalized. If our strategy for measuring prescience is valid, it should rate these 33 individuals as being more prescient than other Republicans of a comparable vintage. As Figure 3 indicates, this is indeed the case.

Next, we consider the relationship between prescience and success in a given domain. If our measure is indeed capturing successful ideas before their success, then first-moving actors who express these ideas should, on average, be recognized and rewarded by the relevant audiences in their field. Consistent with this expectation (Figure 4), prescience is positively related to: a politician's likelihood of being reelected and her status attainment (Panels A and B; SI Table S1); a legal ruling's total citations and probability of being a landmark ruling (Panels C and D; SI Table S2); and a firm's annual stock returns (Panel E; SI Table S3). Indeed, it is only the highly prescient firms (top 5%) that achieve breakthrough levels of cumulative returns (Panel F).

---*Insert Figure 4 Page Width Here*---

The Origins of Prescient Ideas

We turn next to investigating the sources of prescient ideas. Positions in a given domain can be thought of as varying along a continuum from the core, which tends to be occupied by established actors that shape the institutions and norms of a field, to the periphery, which is typically populated by upstart actors that have less ability to coordinate or influence behavior.[15,27] Relative to core actors, those in the periphery typically face fewer institutional constraints and are therefore more likely to generate novel ideas.[17] Although the core/periphery distinction has been widely invoked, it has been operationalized in different ways across empirical contexts. Given that our three empirical settings are quite different from one another, we similarly develop measures of core versus peripheral positions that are specific to and appropriate for each context. These choices are discussed in greater detail in the supplementary information.

Figure 5 shows that, across all three settings, highly prescient ideas—those at or above the 95th percentile of our continuous measure of prescient—emanate from the periphery rather than the core. In politics, eigenvector network centrality (Panel A), K-core network centrality (Panel B), degree centrality, and closeness centrality are negatively related to the likelihood of a politician demonstrating elite prescience (SI Table S5). In law, lower (more peripheral) courts are more likely to produce highly prescient decisions than are upper (more belonging to the core) courts (Panel C; SI Table S6). Indeed, the U.S. Supreme Court, despite having the highest status and the power to set precedent, exhibits the lowest rate of highly prescient decisions. Highly prescient decisions are

22 times more likely to come from the State Appeals Courts than the U.S. Supreme Court. Estimating models with judge fixed effects that take advantage of the fact that judges are sometimes promoted across the judicial status hierarchy, we find that the likelihood of a judge authoring a highly prescient ruling declines by 0.8 percentage points after she is promoted from the U.S. District Court—the lowest rung of the federal judicial hierarchy—to the U.S. Appeals Court.[1] In business, as firm size—a proxy for being part of the core—based on total assets (Panel E) and the number of employees (Panel F) increases, the likelihood of a firm demonstrating elite prescience declines precipitously (SI Table, S7).

---*Insert Figure 5 Page Width Here*---

Discussion

The ability to systematically identify the linguistic markers of prescient ideas enables us to shed light on longstanding questions about the origins of transformative ideas. Popular intuitions often suggest that prescient ideas emerge from powerful incumbents. In law, for example, higher court judges are typically thought to produce prescient rulings which guide the subsequent judgments of lower-court judges.[18] Similarly, in politics, established legislators who lead the most central committees are frequently identified as the most prescient.[28] In business, by contrast, theories of disruptive innovation implicitly assume that the prescient ideas underpinning revolutionary products and business models arise from new entrants to an industry rather than from entrenched incumbents.[29,3] Because we have heretofore lacked a systematic way of quantifying prescience, such intuitions have been mostly informed by anecdotes and case studies. In contrast, our method reveals that, across a diverse array of domains, prescient ideas emanate from the periphery rather than the center.

Our results also suggest that prescience may be as much a property of contexts as of individuals. Indeed, in the legal domain, the same individual becomes less prescient as she moves up the status hierarchy to upper-level courts. Those in search of breakthrough ideas should therefore look beyond the usual suspects who are ensconced in well-trodden places and instead focus attention on the unconventional ideas brewing in the outskirts of a domain.

[1] Given the small number of justices in our data in the U.S. Supreme Court, our estimate is noisier and statistically indistinguishable from the probabilities for judges in U.S. District and Appeals courts.

A pre-condition for an idea to become prescient is that it is contextually novel, yet most novel ideas fail to gain traction.[19-21] Thus, it remains unclear from our study whether a strategy of actively pursuing contextual novelty would lead to long-term success for an actor. We leave to future research the task of unpacking the conditions under which contextual novelty translates into prescience and eventual acclaim in a given field.

Although our empirical investigation focuses on three specific contexts, the method we introduce can detect prescient ideas in other domains. A natural extension, for example, would be the domain of science, where our empirical strategy could be readily applied to the text of academic papers and patents. Indeed, prior work in the "science of science" tradition has shown that disruptive innovation, as defined by citation networks, tends to originate in small and peripheral scientific teams rather than large and core ones.[6,30,3] It remains to be explored the conditions under which prescient ideas, as determined by our measurement approach, translate into disruptive innovation as measured by citations to the tangible artifacts of scientific production. By focusing attention on and providing a novel means to quantify prescient ideas, we aim to broaden scholarly exploration from a narrow fixation on outcomes such as citations to the larger process by which prescient ideas that change the world emerge.

Acknowledgments

We thank Yixi Chen for her invaluable research assistance in collecting and processing the data. We also thank Dan McFarland and Dashun Wang for their helpful comments on prior drafts. The usual disclaimer applies.

Supplementary Material

Supplementary material is included in submission.

Author contributions statement

- Conceptualization: P.V., A.G., and S.B.V.
- Data curation: P.V., A.G., and S.B.V.
- Methodology: P.V.
- Investigation: P.V., A.G., and S.B.V.

- Visualization: P.V.
- Writing - Original Draft: P.V., A.G., and S.B.V.
- Writing - Review & Editing: P.V., A.G., and S.B.V.

Data availability

The measures of prescience and code used to fine-tune BERT models and compute prescience is provided on the first author's GitHub page. Other data, including the raw text, dependent variables, and controls, may be found at separate repositories referenced in this manuscript.

Materials and Methods

To extract prescience from conversational text data, we build upon a recent innovation in natural language processing and deep learning: Bidirectional Encoder Representations from Transformers (BERT).[14] BERT is a generalized language model, meaning it learns the syntax and semantic meaning of language, which can then be applied to a litany of downstream tasks like machine translation and named entity recognition. Underlying BERT is layers of transformer blocks. The transformer architecture diverges from previous language modeling approaches by replacing recurrence and convolutions with attention mechanisms.[31] Doing so allows the entire sentence to be propagated through the model simultaneously, significantly speeding up parallelization. BERT stacks dozens of transformer blocks, encompassing hundreds of millions of parameters. As a result, BERT learns syntax relationships, semantic meanings, co-references, and even encodes entire syntax trees.[32,33] Because BERT is computationally intensive—often requiring several weeks of time on dedicated cloud tensor processing units (TPUs) to train on a new corpus—researchers typically begin with the pre-trained model provided by Google (where BERT was developed) and fine-tune this model to their own corpora. Through the fine-tuning process, the general meanings learned by BERT can be contextualized to the researchers' specific domains of interest.[34]

Traditional language models process sentences left-to-right, one word at a time, and estimate the conditional likelihood of a word: $(w_i|w_0,w_1,...w_{i-1})$. BERT instead favors bidirectionality—that is, it attends to both the left and right contexts simultaneously.[14] To circumvent the unidirectional constraint, BERT is trained (in part) using a masked language model (MLM) task: 15 percent of the words in a sentence are randomly masked, and the model is tasked to predict the masked tokens.

Sentence (s): Earnings are up this quarter.

Masked s: Earnings are [MASK] this quarter.

The MLM objective differs from “true” language models in that the likelihood of the model generating a sentence is undefined. As a proxy, we use the model’s ability to solve the MLM for each word in the sequence, leaving all other words unmasked. Here, the likelihood of a word is conditional on both the left and right contexts: $(w_i|w_0,w_1,...w_{i-1},w_{i+1},w_{i+2},...)$. While most researchers take the generalized language model features of BERT and add an additional layer on top to solve a downstream task, we instead directly use BERT’s probabilistic modeling of language via MLM to quantify prescience.

Specifically, we task the model to minimize the cross-entropy loss between the predicted and the actual word. Let $\overrightarrow{y_i}$ represent a location in a vector of length N, where N refers to the number of words in the corpus, for word i. This one-hot encoded vector takes a value of one at the index of the masked token and zero otherwise. $\overrightarrow{\hat{y}_i}$, also of length N, is the vector predicted token likelihood obtained through a softmax activation layer predicting the masked token by the BERT model. Model accuracy is evaluated using cross-entropy loss, $-\overrightarrow{y_i} \cdot log\left(\overrightarrow{\hat{y}_i}\right)$. To obtain word-level perplexity, PP_i, which is the inverse-likelihood of the model generating the word, we exponentiate the cross-entropy loss (Equation 1).

$$PP_i = exp\left(-\overrightarrow{y_i} \cdot log\left(\overrightarrow{\hat{y}_i}\right)\right) \qquad 1$$

Words that are trivially predicted by the model—such as stop words and punctuation—when used in expected contexts have perplexities of approximating 1, meaning that the model predicts them with close to 100 percent accuracy. Conversely, words and phrases that are highly unusual or unexpected have higher perplexity scores. We take the product of these perplexities and normalize by the nth root to account for sentence length (Eq. 2).

$$CN(s) = \left(\prod_{i=0}^{N} PP_i\right)^{\frac{1}{N}} \qquad 2$$

We refer to this term as contextual novelty (CN) instead of sentence-level perplexity for two reasons. First, given we use this measure to assess the extent to which ideas rethink the contextual assumptions in a domain, terming it contextual novelty aligns our empirical measure with our theoretical quantity of interest. Second, because BERT models bidirectionally, the perplexity of the sentence is technically undefined, and terming this sentence-level perplexity would be inconsistent with prior work.[35]

To measure *prescience*, we rely on the intuition that prescient ideas depart from prevailing ideas at the time of introduction but become commonplace in the future.[36,3,37] For example, Gerow et al.[23] identify highly influential scholarly publications by studying how academic discourse shifts after their publication. Thus, rather than fine-tuning BERT to our entire corpus, we split the corpus into time periods and fine-tune separate BERT models one each split of the corpus. This approach allows us to examine how the contextual novelty of a sentence changes over time.

Our method requires two BERT models trained on a corpus split into two periods, current (t_0) and future (t_1), and two BERT models which map documents to contextual novelties $CN_{t0}(s)$, $CN_{t1}(s)$. For a document from the current period, we define *prescience*, P, as the percentage reduction in contextual novelty between the current and future models.

$$P(s) = \frac{CN_{t0}(s) - CN_{t1}(s)}{CN_{t0}(s)} \qquad 3$$

Both *contextual novelty* and *prescience* are defined at the sentence level. To transition from the sentence level to the relevant unit of analysis in a given domain, we simply take the mean value of these variables over the unit of aggregation. For our corpus of U.S. State and Federal judicial decisions, for example, we aggregate at the unit of the judicial decision. We define truly prescient ideas (as manifested in individuals, organizations, or documents) as those in the top five percent of the distribution of prescience. We find that our measure is noisier with short sentences, as there is less context for BERT to use when making predictions. To reduce noise, we restrict to sentences with at least 10 tokens and less than 100 tokens (to catch errors in the sentence parser) before aggregating to mean prescience. Researchers replicating this methodology may consider using a higher minimum token count, such as 15 or 20 tokens, to further reduce noise.

Data, Fine-Tuning, and Measuring Vision

Training BERT from scratch is prohibitively expensive, taking weeks on a cloud TPU. Instead, Google has provided a pre-trained model—trained on the BookCorpus (800M tokens) and the English Wikipedia (2.5B tokens) available at https://github.com/google-research/bert—that researchers can fine-tune on their own corpora to learn context-specific idiosyncrasies. We fine-tune using BERT-Base uncased (12-layer, 768-hidden, 12-head, 110M parameter model) by repeating the MLM task and next-sentence prediction task on our corpora. We filter out sentences with less than 10 tokens to reduce noise and sentences longer than 100 tokens given that they likely represent errors in the sentence parser[2]. For fine-tuning, we use a max sequence length of 128, a batch size of 64, and fine-tune for approximately 400,000 steps. The only pre-processing of text prior to fine-tuning is sentence tokenization and appending [CLS] and [SEP] tokens to the start and end of each sentence respectively. BERT uses WordPeiece tokenization, which converts unrecognized tokens into sub-tokens (e.g., tokenizing onboarding into [onboard, ing] so no out-of-vocabulary words are dropped from the analysis.

Below we describe the three data sets in greater detail and explain the specific text pre-processing, fine-tuning, and approach to computing prescience we followed in each setting. For a full description please refer to the supplemental information.

Politics

To identify prescient politicians, we use transcripts from the United States House of Representatives and the United States Senate from the bound and daily editions of the United States Congressional Record from the 43rd to 114th Congresses (1873-2017). We use floor speeches, as opposed to other political texts such as press releases or campaign speeches, as these texts 1) maintain a consistent format and 2) offer a complete population of political discourse. We use the data set provided by Gentzkow, Shapiro, and Taddy (2019), who remove procedural language and parse the text from each congressional session into speeches attributable to congresspeople.[38] For data reliability reasons (e.g., temporal variation in optical character recognition (OCR) accuracy), we begin our analyses with the 87th Congress, whose members took office in 1961 resulting in 4.9 million unique speech events. We focus on this later subset of the corpus because it contains fewer OCR errors, which can potentially bias our measure of prescience. We obtain biographical data on politicians from the Congressional Biographical Directory, GovTrack, and Congress.gov and collect committee membership[39,40] and bill cosponsorship data.[41]

We segment the corpus into four-year increments, corresponding to presidential terms, resulting in 15 buckets—19611964, 1965-1968, ... 2007-2010—and fine-tune separate BERT models for each increment. We compartmentalize corpus by presidential terms given that these are natural breakpoints in the trajectory of political discourse. To compute *prescience*, we define the current period as the BERT model trained using the year the sentence was spoken, M_0. A more difficult

[2] An even higher minimum token count will greatly reduce noise in computing contextual novelty and prescience.

[3] Empirically, the prescience calculations between the two models are highly correlated, $\rho \approx 0.7$.

choice is in selecting the future period. Choosing a proximate model quantifies short-term evolution in discourse, while a model trained on text further in time from the focal sentence quantifies longer-term prescience. To balance this trade-off, we select two future period models—the immediately subsequent model M_1 and the model after that M_2—and take the arithmetic mean of prescience between these two prescience calculations: $[P(M_0,M_1) + P(M_0,M_2)]/2$. We define the current period model as the BERT model trained using the year the sentence was spoken. So, for example, for a sentence spoken in 1965, we use the BERT model trained on sentences from 1965 to 1968 as M_0 and the 1969-1972 and 1973-1976 models as the future models.[3]

Law

Our data of U.S. State and Federal cases comes from the Caselaw Access Project, which digitized and processed over eight million state and federal judicial verdicts, stretching back to the 1640s. These data provide the complete population of legal discourse and are the main forum of judicial interpretation and legal precedent. As with our data on political discourse, we found significantly more OCR errors prior to 1960. Thus, to both align these data with data on politicians and to minimize OCR errors, we restrict our analysis to cases beginning in 1960. Our resulting sample includes 4.2 million cases. We obtain biographical data on federal judges, including their court tenure, gender, and prior judicial service from the Federal Judicial Center. Data on case citations come from the citation graph provided by the Caselaw Access Project, which extracts citations from the in-line text of court decisions. We remove in-line citations using LexNLP, a python package specifically designed to parse legal text, and sentence tokenize using this package as well.

We compartmentalize the corpus into five-year intervals—1960-1964, 1965-1969...2005-2009—and fine-tune separate BERT models on each interval. As with the politics data set, we define the current period model as the BERT model trained using the focal year and compute mean prescience using the subsequent two models to strike a balance between assessing shorter-term versus longer-term prescience.

Business

To study prescient ideas in business, we collect a corpus of quarterly earnings calls (QECs) from seekingAlpha, a content service for financial markets. Our data set includes 108,334 QECs (414 million tokens) from publicly traded firms, predominately headquartered in the United States from 2006 and 2016. We restrict our analyses to the Q&A section of the call, removing the prepared remarks by the company and filtering out statements by analysts and the operator. We restrict to the Q&A portion because prepared statements have a very different style of discourse than the Q&A section which may complicate fine-tuning on a smaller corpus. The back-and-forth discussion in QECs, unlike alternative forms of corporate text such as annual reports or press releases, offers intimate insights into the potentially prescient ideas of firms. We disambiguate company names in quarterly earnings calls and fuzzy match them to Compustat to obtain firm characteristics and performance outcomes. We identify 5,847 firms that have corresponding links to Compustat gvkeys. We then match gvkeys with Permno to link to the CRSP database, thereby allowing us to collect data on daily stock returns.

Unlike our datasets of political speeches and legal decisions, this corpus is comparatively small. We lack a significant number of transcripts for speeches between 2006 and 2011 and the text in the earlier speeches is heavily influenced by the 2007/2008 financial crisis. As a result, we restrict our analysis to QECs from 2011 onward. Given that these data span a relatively small number of years, we split the corpus on an annual basis and fine-tune a separate BERT model for each year in our data. We select 2011 as the focal year and select both 2015 and 2016, the last two years in our data, as the future comparison periods when computing prescience. Because the unit of analysis is the firm, we aggregate all QECs for each firm in 2011 by computing the mean prescience across calls in a given year to create a firm-level measure of prescience.

Validating Vision

In Figure 4 of the main manuscript, we demonstrate how prescient actors reap rewards. For more details on these analyses and their associated regression tables please refer to the SI.

Politics

We begin by considering the extent to which prescience appears to be rewarded for politicians in our data. We aggregate political prescience at the congressional-term unit of analysis and select two measures of political success. The first, political *reelection*, is widely considered the primary motivator of incumbent politicians.[42] The second variable, *committee status*, measures political success within the legislative chamber itself. We use the transfer ratio defined by Bullock and Spraque (1969), which is computed for each committee.[43] We estimate linear regressions of reelection and committee status on average prescience (Table S1).

Law

We turn next to considering whether prescience is associated with success in the legal domain. We define success as the number of legal citations that accrue to the case. We restrict our analyses of citations to federal cases because doing so enables us to include judge-level controls, such as judge fixed effects, which are unavailable for state-level decisions. We also restrict to cases with at least 50 sentences to reduce variance in prescience. This restriction drops less than one percent of decisions.

The first dependent variable of interest, *log citations*, is the natural log of the number of citations a case receives plus one. We log transform the variable to reduce skew. A more stringent test of our method is whether it can identify landmark decisions—judicial rulings that significantly alter existing interpretations of existing law. Most landmark decisions come from the U.S. Supreme Court, given they have absolute authority to set precedents and determine law. We define a U.S. Supreme Court ruling as a *landmark decision* if it is above the 95*th* percentile in citations for the court. We estimate *log citations* using linear regression and *landmark decision* using logistic regression (Table S2).

Business

To validate our measure of prescience in the business context, we began by testing whether our measure of prescience can predict future stock returns. We measure stock returns using daily returns data: the percentage change in value accrued to the stockholder on a given day. We report the model looking at three-year returns from 2012 to 2014 in Figure 4, Panel E, although our findings are robust to all windows (Table S5). We use 3-digit North American Industry Classification System (NAICS) fixed effects in all regressions to adjust for industry-level heterogeneity (Table S3 & S4).

Prescience Arises from the Periphery of a Field

Next, we turn to the core claim of the paper: prescient ideas are more likely to emerge from the periphery than from the center. There are many different ways to operationalize core versus peripheral positions—for example, based on an actor's position in the network of actors, relative size, and social status (e.g., based on demographic traits such as gender or elite credentialing). We use each of these different approaches in our three empirical settings. The choice of how to operationalize core versus peripheral positions is based on our understanding of each empirical context and the nature of the data that was available to us. The following section describes in detail the data and analyses used to create Figure 5 of the main paper.

We define highly prescient politicians, judicial decisions, and firms as the top five percent in prescience during their relevant measurement period. For politicians, this means the top five percent in their current legislative term. We select the 95*th* percentile and above as prior work predicting scientific impact uses this threshold.[3,20] When predicting *highly prescient*, we control for the number of sentences for each actor, because those with relatively few utterances are disproportionately likely to exhibit high levels of prescience simply because our measure of vision is noisier with fewer data. We do so using *log sentence count*.

Politics

We define core and peripheral positions in the legislative body via the politician's network position, as prior research demonstrates that well-connected politicians can influence their peers and governmental policy.[41] To reconstruct the network of relationships between politicians we use data on bill cosponsorships[44,45,46] and construct separate bill consponsorship networks for each congressional term. We use a variety of network statistics to measure how peripheral a politician is in the cosponsorship network, including *degree centrality*, *eigenvector centrality*, *closeness centrality*, and *K-core centrality*. Because network centrality measures vary over time and between chambers, we standardize each network measure at the congressional term x chamber level. We regress *highly prescient politician* on our suite of network covariates using logistic regression (Table S5).

Law

In our legal setting, the U.S. Supreme Court is the most firmly in the core because it has statutory authority over all other courts. Recognizing that jurisdictional differences render it difficult to make direct comparisons, we assume that state courts are more peripheral than federal courts because the fragmented body of state laws is consolidated by federal rulings but not vice versa. Thus, our judicial hierarchy goes from most peripheral to most core in the following order: State Appeals, State Supreme, U.S. District, U.S. Appeals, U.S. Supreme. We define *highly prescient decision* as the top five percent most prescient decisions in a given year across all courts. To test the idea that prescient judicial decisions come from the structural periphery of the court system, we regress *highly prescient decision* on separate indicator variables for each court using logistic regression (Table S6).

Business

Whereas the periphery in the legal system is defined by hierarchy, we define the periphery in business based on firm size. Large firms command significant market power and use their power to lobby congress for favorable regulations, set industry standards, and influence prices. By virtue of their size, large firms also occupy more central positions in the network between firms.[47] We define firm size using *log total assets*, the total amount of economic value held by the firm and its number of employees, operationalized as *log employee count*. We define a *highly prescient firm* as one at or above the 95*th* percentile in vision and estimate using logistic regression (Table S7).

References

[1] Walter Isaacson. *Steve Jobs: The Exclusive Biography*. Simon & Schuster, 2011.

[2] Andrew Roberts. *Napoleon: A Life*. Penguin Books, reprint edition, 2015.

[3] Lingfei Wu, Dashun Wang, and James A. Evans. Large teams Develop and Small Teams Disrupt Science and Technology. *Nature*, 566(7744):378–382, 2019.

[4] Jasjit Singh and Lee Fleming. Lone Inventors as Sources of Breakthroughs: Myth or Reality? *Management Science*, 56(1):41–56, 2010.

[5] Bas Hofstra, Vivek V. Kulkarni, Sebastian Munoz-Najar Galvez, Bryan He, Dan Jurafsky, and Daniel A. McFarland. The Diversity–Innovation Paradox in Science. *Proceedings of the National Academy of Sciences*, 117(17):9284– 9291, 2020.

[6] Stefan Wuchty, Benjamin F. Jones, and Brian Uzzi. The Increasing Dominance of Teams in Production of Knowledge. *Science*, 316(5827):1036–1039, 2007.

[7] Yifang Ma and Brian Uzzi. Scientific prize network predicts who pushes the boundaries of science. *Proceedings of the National Academy of Sciences*, 115(50):12608–12615, 2018.

[8] Jesse N. Curtis. Remembering Racial Progress, Forgetting White Resistance: The Death of Mississippi Senator John C. Stennis and the Consolidation of the Colorblind Consensus. *History and Memory*, 29(1):134–160, 2017.

[9] Thomas S. Kuhn. *The Structure of Scientific Revolutions: 50th Anniversary Edition*. University of Chicago Press, 2012.

[10] Nikhil Garg, Londa Schiebinger, Dan Jurafsky, and James Zou. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115(16):E3635–E3644, 2018.

[11] Gabriel Grand, Idan Asher Blank, Francisco Pereira, and Evelina Fedorenko. Semantic projection recovers rich human knowledge of multiple object features from word embeddings. *Nature Human Behaviour*, 2022.

[12] Pedro L. Rodriguez and Arthur Spirling. Word embeddings: What works, what doesn’t, and how to tell the difference for applied research. *The Journal of Politics*, 84(1):101–115, 2022.

[13] Austin C. Kozlowski, Matt Taddy, and James A. Evans. The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review*, 84(5):905–949, 2022.

[14] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186, 2019.

[15] Robert K. Merton. *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press, 1973.

[16] Stephen P Borgatti and Martin G Everett. Models of core/periphery structures. *Social networks*, 21(4):375–395, 2000.

[17] Gino Cattani and Simone Ferriani. A Core/Periphery Perspective on Individual Creative Performance: Social Networks and Cinematic Achievements in the Hollywood Film Industry. *Organization Science*, 19(6):824–844, 2008.

[18] David A. Schultz and Christopher E. Smith. *The Jurisprudential Vision of Justice Antonin Scalia*. Rowman & Littlefield, 1996.

[19] Lee Fleming. Recombinant Uncertainty in Technological Search. *Management Science*, 47(1):117–132, 2001.

[20] Brian Uzzi, Satyam Mukherjee, Michael Stringer, and Ben Jones. Atypical Combinations and Scientific Impact. *Science*, 342(6157):468–472, 2013.

[21] Jacob G. Foster, Andrey Rzhetsky, and James A. Evans. Tradition and Innovation in Scientists’ Research Strategies. *American Sociological Review*, 80(5):875–908, 2015.

[22] Lee Coppola and Nicholas DeMarco. Civil RICO: How ambiguity allowed the racketeer influenced and corrupt organizations act to expand beyond its intended purpose. *New England Journal on Criminal and Civil Confinement*, 38(2):241–256, 2012.

[23] Aaron Gerow, Yuening Hu, Jordan Boyd-Graber, David M. Blei, and James A. Evans. Measuring discursive influence across scholarship. *Proceedings of the National Academy of Sciences*, 115(13):3308–3313, 2018.

[24] Sam Arts, Jianan Hou, and Juan Carlos Gomez. Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures. *Research Policy*, 50(2):104144, 2021.

[25] Robert Mann. *Legacy to Power: Senator Russell Long of Louisiana*. Paragon House, 1992.

[26] Sean M. Theriault. *The Gingrich Senators: The Roots of Partisan Warfare in Congress*. Oxford University Press, 2013.

[27] David Knoke, Franz Urban Pappi, Jeffrey Broadbent, and Yutaka Tsujinaka. *Comparing Policy Networks: Labor Politics in the US, Germany, and Japan*. Cambridge University Press, 1996.

[28] Dan Diller and Sara Stefani. *Richard G. Lugar: Indiana's Visionary Statesman*. Indiana University Press, 2019.

[29] Clayton Christensen, Rory Morgan McDonald, Elizabeth J. Altman, and Jonathan Palmer. Disruptive Innovation: Intellectual History and Future Paths. *Academy of Management Proceedings*, 2017, 2017.

[30] Benjamin F. Jones, Stefan Wuchty, and Brian Uzzi. Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science. *Science*, 322(5905):1259–1262, 2008.

[31] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In *31st Conference on Neural Information Processing Systems*, 2017.

[32] John Hewitt and Christopher D Manning. A Structural Probe for Finding Syntax in Word Representations. In *Proceedings of NAACL-HLT*, pages 4129–4138, 2019.

[33] Kevin Clark, Urvashi Khandelwal, Omer Levy, and Christopher D. Manning. What Does BERT Look at? An Analysis of BERT's Attention. In *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 276–286, 2019.

[34] Yoshua Bengio. Deep Learning of Representations for Unsupervised and Transfer Learning. In *Proceedings of ICML workshop on unsupervised and transfer learning*, pages 17–37, 2012.

[35] R. Rosenfeld. Two decades of statistical language modeling: where do we go from here? *Proceedings of the IEEE*, 88(8):1270–1278, 2000.

[36] Bryan Kelly, Dimitris Papanikolaou, Amit Seru, and Matt Taddy. Measuring Technological Innovation over the Long Run. Working Paper 25266, National Bureau of Economic Research, 2020.

[37] Russell J. Funk and Jason Owen-Smith. A Dynamic Network Measure of Technological Change. *Management Science*, 63(3):791–817, 2016.

[38] Matthew Gentzkow, Jesse M. Shapiro, and Matt Taddy. Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech. *Econometrica*, 87(4):1307–1340, 2019.

[39] Charles Stewart III. Committee Hierarchies in the Modernizing House, 1875-1947. *American Journal of Political Science*, 36(4):835–856, 1992.

[40] Charles Stewart III and Jonathan Woon. Congressional Committee Assignments, 103rd to 114th Congresses, 1993-2017, 2017.

[41] James H. Fowler. Connecting the Congress: A Study of Cosponsorship Networks. *Political Analysis*, 14(4):456–487, 2006.

[42] Joseph A. Schlesinger. *Ambition and Politics: Political Careers in the United States*. Rand McNally and Co, Chicago, IL, 1966.

[43] Charles Bullock and John Sprague. A Research Note on the Committee Reassignments of Southern Democratic Congressmen. *The Journal of Politics*, 31(2):493–512, 1969.

[44] James H. Fowler. Legislative cosponsorship networks in the US House and Senate. *Social Networks*, 28(4):454–465, 2006.

[45] Leanne Ten Brinke, Christopher C Liu, Dacher Keltner, and Sameer B Srivastava. Virtues, vices, and political influence in the us senate. *Psychological Science*, 27(1):85–93, 2016.

[46] Christopher C Liu and Sameer B Srivastava. Efficacy or rigidity? power, influence, and social learning in the us senate, 1973–2005. *Academy of Management Discoveries*, 5(3):251–265, 2019.

[47] Michael Patrick Allen. The Structure of Interorganizational Elite Cooptation: Interlocking Corporate Directorates. *American Sociological Review*, 39(3):393–406, 1974.

Figures

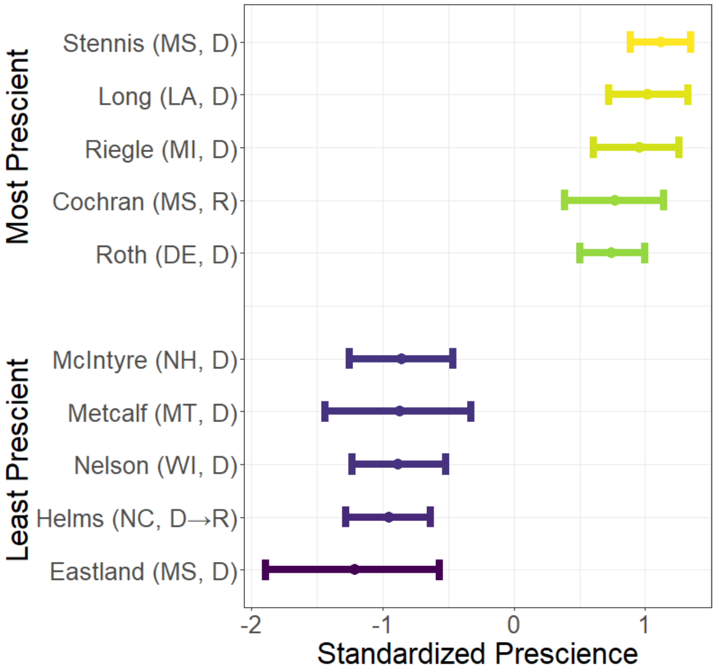

Figure 1: Prescient Senators, minimum three congressional terms. Mean prescience is computed, standardized, and bootstrapped 10K times at the politician-quarter level.

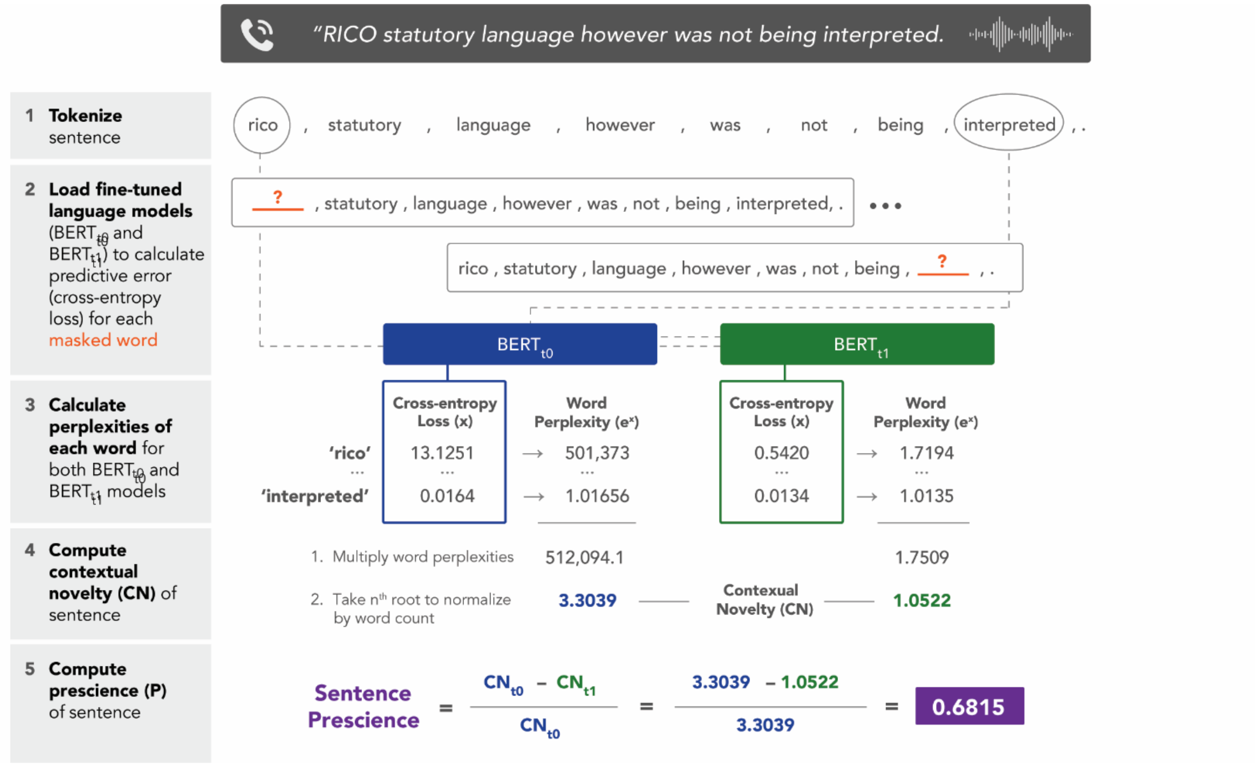
"RICO statutory language however was not being interpreted.
1 Tokenize sentence
2 Load fine-tuned language models (BERT$_{t0}$ and BERT$_{t1}$) to calculate predictive error (cross-entropy loss) for each masked word
3 Calculate perplexities of each word for both BERT$_{t0}$ and BERT$_{t1}$ models
4 Compute contextual novelty (CN) of sentence
5 Compute prescience (P) of sentence
rico , statutory , language , however , was , not , being , interpreted , .
? , statutory , language , however , was , not , being , interpreted, .
• • •
rico , statutory , language , however , was , not , being , ? , .
BERT$_{t0}$
BERT$_{t1}$
Cross-entropy Loss (x)
Word Perplexity (e^x)
'rico'
'interpreted'
13.1251 → 501,373
0.0164 → 1.01656
0.5420 → 1.7194
0.0134 → 1.0135
1. Multiply word perplexities
512,094.1
1.7509
2. Take nth root to normalize by word count
3.3039
Contexual Novelty (CN)
1.0522
Sentence Prescience = $\frac{CN_{t0} - CN_{t1}}{CN_{t0}}$ = $\frac{3.3039 - 1.0522}{3.3039}$ = 0.6815

Figure 2: Illustration of how prescience is computed based on a sentence from the legal dataset that the model deems highly prescient. This sentence rates highly in prescience because the RICO (Racketeer Influenced and Corrupt Organizations Act) token is better predicted the future period, when RICO's statutory language was heavily contested by the courts.[9]

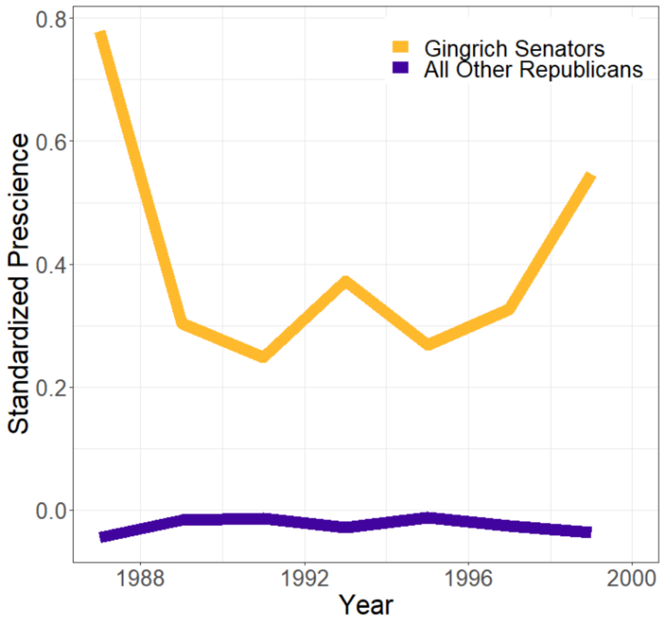

Figure 3: Gingrich senators. Average standardized prescience for the Gingrich senators and all other Republicans by congressional term.

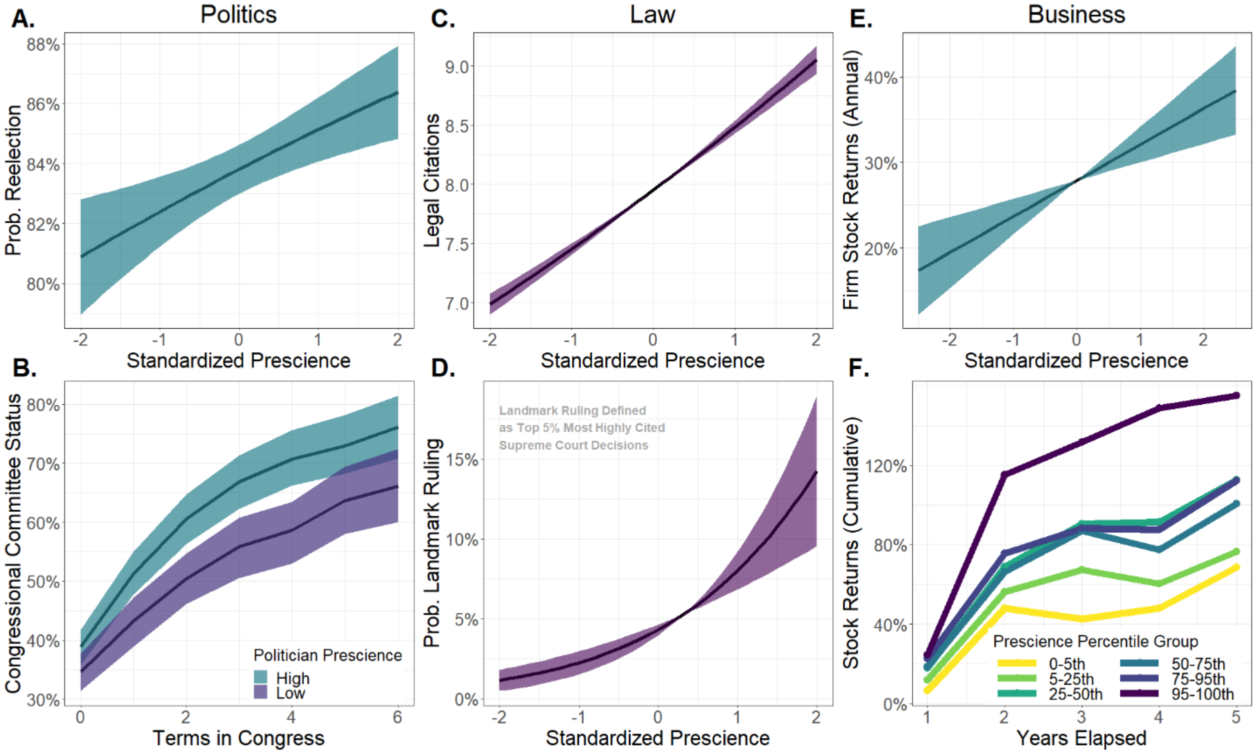

Figure 4: Prescience Predicts Success a) Prescience predicts political re-election. Marginal effects plot from panel linear probability models of political reelection on standardized prescience; politician-term unit of analysis, with political party x congressional term fixed effects (β = 0.00860, p < 0.05). b) Prescience predicts congressional committee status. Mean committee status for the top tercile (high prescience) and bottom tercile (low prescience) Congressional term with 10K politician bootstrapped SEs. Congressional committee status defined by the committee transfer ratio (SI Appendix). Please see SI for panel regressions with fixed effects and other controls (β =0.0155, p < 0.001). c) Prescience predicts highly cited court decisions. Marginal effects plot of linear regression model of log total citations on standardized prescience; judicial decision unit of analysis, with judge, court, and year fixed effects (β =0.0693, p < 0.001). d) Prescience predicts landmark Supreme Court decisions. Marginal effects plot of linear probability models of landmark decisions on standardized prescience. Landmark decision is defined as the top 5% most highly cited U.S. Supreme Court decisions by year, with the sample restricted to U.S. Supreme Court decisions. Models include judge and year fixed effects (β =0.687, p < 0.001). e) Prescience predicts firm stock returns. Marginal effects plot of linear regression models of yearly stock returns from 2012-2015 on 2011 standardized prescience; NAICS 3-digit industry fixed effects (β =0.0422, p < 0.001). f) Prescience predicts elite firm performance. Total stock returns since 2012 by prescience quartile and year (with top and bottom 5%). The y-axis shifts from annual stock returns in panel e to cumulative stock returns in panel f.

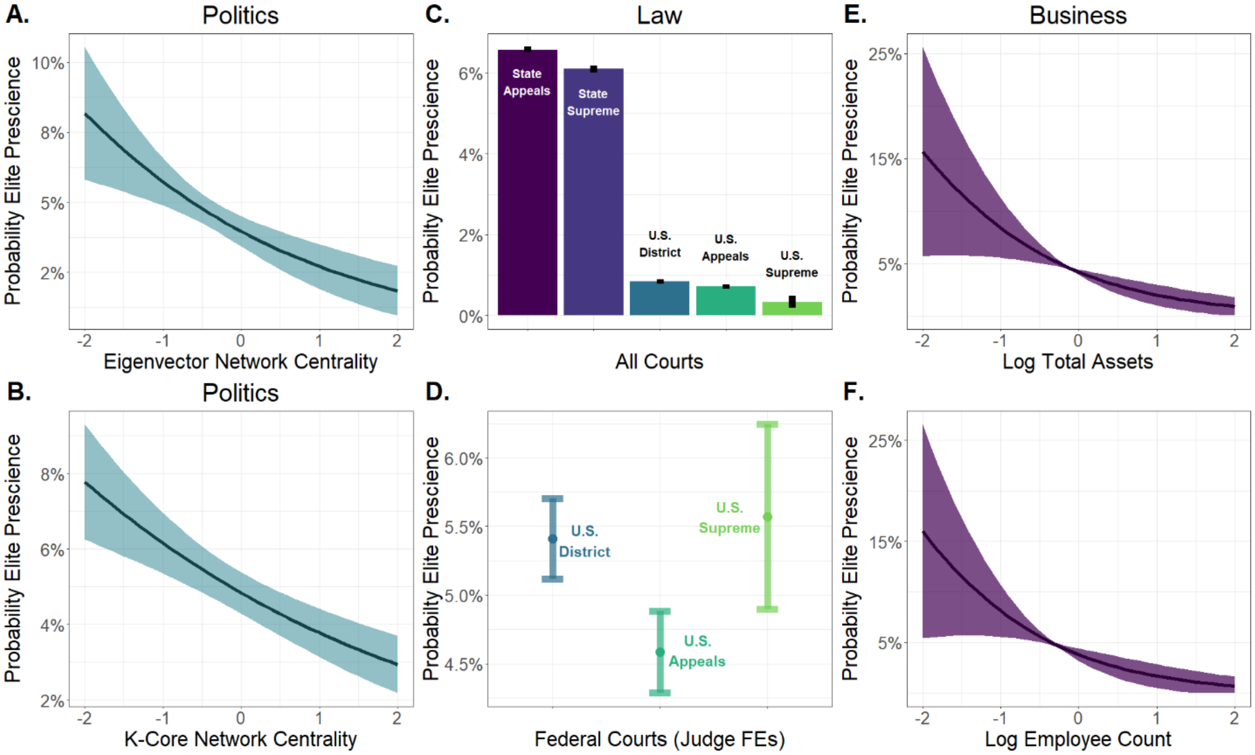

Figure 5: Highly prescient ideas come from the periphery. Marginal effects plots regressing the probability of having elite prescience (top 5% in standardized prescience) on alternatives measures of peripheral positions using logistic regression and 95% confidence intervals. All regressions include controls for the log number of sentences and are restricted to observations with at least 50 sentences given increased variance in prescience with small sample size. a & b) Highly prescient politicians come from peripheral network positions. Politician network defined using bill consponsorship data.[15] Network periphery measured by standardized eigenvector centrality ($\beta = -0.292$, $p < 0.001$) and standardized k-core centrality ($\beta = -0.277$, $p < 0.001$) with additional centrality measures in the SI. c & d) Highly prescient court decisions come from the lower courts. Panel C depicts the probability of a prescient decision using both state and federal courts and year fixed effects. Panel D adds judge fixed effects and restricts the sample to federal decisions (for which we have judge disambiguated decisions). e & f) Highly prescient ideas come from small firms. NAICS 2-digit industry fixed effects. Firm size measured by standardized total assets ($\beta = -.207$, $p < 0.01$) and the standardized number of employees ($\beta = -0.309$, $p < 0.05$).